WHY MULTI-SOURCE, MULTI-TARGET DATA FUSION IS TRICKY

May 7, 1999

Ronald P.S. Mahler*
Lockheed Martin Tactical Defense Systems
3333 Pilot Knob Rd, Eagan MN 55121
TEL: 651-456-4819; EML: ronald.p.mahler@lmco.com

ABSTRACT

The purpose of this paper is to demonstrate that, when the number of targets is not known beforehand, Bayesian optimal filtering approaches to multisensor-multitarget data fusion problems encounter unexpected conceptual and practical difficulties. The reason is that single-target Bayesian filtering cannot be naïvely generalized to multitarget situations and that, consequently, serious pitfalls await those who simply "declare victory." In particular, we show that the classical Bayesian techniques for optimally determining parameters of interest--e.g., the maximum a posteriori (MAP) and expected a posterior (EAP) estimators--cannot even be defined in multitarget situations. We describe our solution to this problem, "finite-set statistics" (FISST), as well as "joint multitarget probabilities (JMP)," a renaming of a special case of FISST. We show how FISST leads to provably Bayes-optimal multisensor-multitarget data fusion algorithms. We discuss the optimality and convergence properties of two different Bayesian data fusion algorithms.

1 Introduction

Suppose that a single sensor observes a single moving target and that:

(1) $\mathbf{x}_\alpha$ is the state of the target at time-step α;
(2) $\mathbf{z}_\alpha$ is the measurement of the target collected by a single sensor at time-step α;
(3) $Z^\alpha = \{\mathbf{z}_1, \ldots, \mathbf{z}_\alpha\}$ is the set of accumulated evidence at time-step α;
(4) $f(\mathbf{z} \mid \mathbf{x})$ is the sensor likelihood function;
(5) $f_{\alpha+1\mid\alpha}(\mathbf{x}_{\alpha+1} \mid \mathbf{x}_\alpha)$ is the Markov motion model for the target; and
(6) $f_{\alpha\mid\alpha}(\mathbf{x}_\alpha \mid Z^\alpha)$ is the posterior density of the target state at time-step α, given all accumulated data Z^α

Then the following well-known *Bayesian recursive nonlinear filtering equations* are the basis for recursive Bayesian optimal single-target tracking **[1,2,7,9 p. 174,41]**:

$$(7a)\qquad f_{\alpha+1\mid\alpha}(\mathbf{x}_{\alpha+1} \mid Z^\alpha) = \int f_{\alpha+1\mid\alpha}(\mathbf{x}_{\alpha+1} \mid \mathbf{x}_\alpha)\, f_{\alpha\mid\alpha}(\mathbf{x}_\alpha \mid Z^\alpha)\, d\mathbf{x}_\alpha$$

* The work reported in this paper was supported by the U.S. Army Research Office under contracts DAAH04-94-C-0011 and DAAG55-98-C-0039. The content does not necessarily reflect the position or the policy of the Government. No official endorsement should be inferred.

Form SF298 Citation Data

Report Date ("DD MON YYYY")	Report Type	Dates Covered (from... to) ("DD MON YYYY")
07051999	N/A	

Title and Subtitle Why Multi-Source, Multi-Target Data Fusion is Tricky	**Contract or Grant Number**
	Program Element Number
Authors Mahler, Ronald P. S.	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Lockheed Martin Tactical Defense Systems 3333 Pilot Knob Rd. Eagan MN 55121	**Performing Organization Number(s)**
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Monitoring Agency Acronym**
	Monitoring Agency Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes	
Abstract	
Subject Terms	
Document Classification unclassified	**Classification of SF298** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** unlimited
Number of Pages 20	

$$f_{\alpha+1|\alpha+1}(\mathbf{x}_{\alpha+1} | Z^{\alpha+1}) \qquad \frac{f(\mathbf{z}_{\alpha+1} | \mathbf{x}_{\alpha+1})\, f_{\alpha+1|\alpha}(\mathbf{x}_\alpha | Z^\alpha)}{\int f(\mathbf{z}_{\alpha+1} | \mathbf{x}_{\alpha+1})\, f_{\alpha+1|\alpha}(\mathbf{x}_\alpha | Z^\alpha)\, d\mathbf{x}_{\alpha+1}}$$

Given the posterior, the current target state $\mathbf{x}_\alpha$ can be approximated using any of the classical Bayes-optimal estimators, e.g. the maximum *a posteriori* (MAP) or expected *a posteriori* (EAP) estimators.

1.1 "Naïve" Multitarget Recursive Bayesian Nonlinear Filtering. Ideally, one would like to *directly generalize* this standard and well-understood approach to *multisensor, multitarget* problems. That is, suppose that

(1′) $X_\alpha = \{\mathbf{x}_{1;\alpha}, \ldots, \mathbf{x}_{n(\alpha);\alpha}\}$ is the set of states of the unknown number $n(\alpha)$ of targets at time-step α;
(2′) $Z_\alpha = \{\mathbf{z}_{1;\alpha}, \ldots, \mathbf{z}_{k(\alpha);\alpha}\}$ is the set of observations collected by all sensors at or around time-step α;
(3′) $Z^{(\alpha)} : Z_1, \ldots, Z_\alpha$ is the accumulated evidence at time-step α;
(4′) $f(Z \mid X)$ is the likelihood function for the *entire suite of sensors* collecting on all targets;
(5′) $f_{\alpha+1|\alpha}(X_{\alpha+1} \mid X_\alpha)$ is the multitarget Markov motion model for the entire multitarget system; and
(6′) $f_{\alpha|\alpha}(X_\alpha \mid Z^{(\alpha)})$ is the multitarget posterior density of the multitarget state X_α at time-step α, given all evidence $Z^{(\alpha)}$ accumulated up until that time.

As viewed from a "naïve" perspective, it would seem that all that is needed is to write down the following *multitarget* versions of the recursive Bayesian nonlinear filtering equations:

$$(7a') \qquad f_{\alpha+1|\alpha}(X_{\alpha+1} | Z^{(\alpha)}) = \int f_{\alpha+1|\alpha}(X_{\alpha+1} | X_\alpha)\, f_{\alpha|\alpha}(X_\alpha | Z^{(\alpha)})\, \delta X_\alpha$$

$$(7b') \qquad f_{\alpha+1|\alpha+1}(X_{\alpha+1} | Z^{(\alpha+1)}) = \frac{f(Z_{\alpha+1} | X_{\alpha+1})\, f_{\alpha|\alpha}(X_\alpha | Z^{(\alpha)})}{\int f(Z_{\alpha+1} | X_{\alpha+1})\, f_{\alpha|\alpha}(X_\alpha | Z^{(\alpha)})\, \delta X_{\alpha+1}}$$

Likewise, given the multitarget posterior distribution $f_{\alpha|\alpha}(X_\alpha \mid Z^{(\alpha)})$, we would need only apply one of the classical Bayes-optimal estimators to arrive at an estimate of the current multitarget state X_α. We would then be able to estimate, in a *simultaneous and optimal* manner, the number $\hat{n}$ of targets as well as the individual states $\hat{\mathbf{x}}_i$ of the targets, without any need to determine an optimal report-to-track association. Stated in different words: Such a procedure would optimally resolve the conflicting objectives of detection and tracking (as well as identification, if target I.D. state and measurement variables are present).

Unfortunately, when one tries to apply the standard statistical thinking just described to the *multitarget case with unknown number of targets*, one quickly discovers that taking things for granted leads to serious troubles that directly bear on practice. Specifically, *Bayes-optimal multitarget estimation and filtering encounters fundamental conceptual and practical difficulties when the number of targets is unknown*. What is at issue is neither theoretical hair-splitting nor mere mathematical "bookkeeping." Rather, what is actually at stake is *our ability to do Bayes-optimal multitarget filtering and estimation at all and, moreover, our ability to even know what "Bayes-optimal" means in such a context*.

This paper is the third in a series devoted to explaining these difficulties; the first two were "Multisource, multitarget filtering: A unified approach" **[29]** and "Multitarget Markov Motion Models" **[30]**. In these papers and in the book *Mathematics of Data Fusion* **[6]**, we described a theoretically rigorous, systematic, and unified resolution of these and other difficulties based on a special case of random set theory **[3,8,36]** called "finite-set statistics (FISST) **[6,8,21-26,28-35,40]**. We also explained why FISST (or something of equal rigor) is necessary to deal

with the associated problems. In the second paper we elaborated on item (5′) above: The definition and construction of computationally tractable Markov motion models capable of effectively dealing with the dynamicism of real-world problems. The present paper has two purposes: First, to summarize and explain the major difficulties by contrasting true multitarget recursive Bayesian nonlinear filtering with three approaches that fail to address these difficulties. Second, to elaborate on item (6′): extracting the information contained in the multitarget posterior so that the associated filter will rapidly and stably converge to the correct answer.

1.2 A Short History of Multitarget Recursive Bayesian Nonlinear Filering. The concept of multitarget recursive Bayesian nonlinear filtering is a relatively new one in the data fusion engineering community. The problems associated with it are greatly alleviated (but not completely eliminated) if we assume that *the number of targets, n, is known*. The history of this "known-*n*" aspect of the problem is summarized in Table I.

TABLE I: Multitarget Recursive Bayesian Nonlinear Filtering: No. of Targets Assumed Known

Date	Author(s)	Theoretical Basis
1987	Washburn [45]	Point Processes
1991	Xie and Evans [46]	Hidden Markov models
1992	Kamen et. al. [10-12,38] *"Symmetric Measurement Equations"*	Multitarget measurement model transformed into nonlinear measurement model
1993	Kastella [14] *"Event-Averaged MLE"*	Multitarget EKF using saddle-point approximation
1995	Krishnamurthy and Evans [18]	Hidden Markov models

To our knowledge the earliest work is due to Washburn, who used point process theory to model both multitarget observations and multitarget states and then constructed a multitarget likelihood function defined in terms of these point processes. (Note: Point process theory is closely related to random set theory, see [3].) Kamen and his associates have attacked the problem by constructing a multitarget measurement model for a Gaussian sensor. Specifically, the single-target measurement models $\mathbf{z}_1 = C\mathbf{x}_1 + \mathbf{w}_1, \ldots, \mathbf{z}_n = C\mathbf{x}_n + \mathbf{w}_n$ are transformed into a conventional (but nonlinear) sensor model $\mathbf{Z} = f(\mathbf{X}) + \mathbf{V}$ where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{Z}$ is a vector whose components are symmetric functions of the measurements. In Kastella's approach, a heuristic multitarget average likelihood $f_{EAMLE}(Z \mid X)$ is defined for a Gaussian sensor in Poisson clutter and incorporated into a conventional Extended Kalman Filter (see Section 4 below).

Table II summarizes the history of multitarget Bayesian recursive nonlinear filtering problem when the number n of targets is *not* known and must be determined along with the individual target states. As already noted, in the unknown-*n* case multitarget recursive Bayesian nonlinear filtering runs afoul of serious conceptual, theoretical, and practical difficulties [29,30]. Only some of the researchers listed in Table II have addressed (or even been aware of) these problems. The earliest work appears to be due to Miller, O'Sullivan, Srivistava, et. al. at Washington University in St. Louis. Their very sophisticated approach requires solution of stochastic diffusion equations on non-Euclidean manifolds. It is also unique in that it is apparently the only approach to deal with *continuous* evolution of the multitarget state. (All other approaches listed in Tables I and II assume discrete state-evolution.) Mahler was apparently the first to deal with the discrete state-evolution case in complete generality (Bethel and Paras assume discrete observation and state variables). Kastella's "joint multitarget probabilities" are just a renaming of a special case of Mahler's approach (see Section 6 below).

TABLE II: Multitarget Recursive Bayesian Nonlinear Filtering: No. of Targets Unknown

Date	Author(s)	Theoretical Basis
1991	Miller et. al. [19,37,42] *"Jump Diffusion"*	Approx. solution of multitarget stochastic PDEs
1994	Bethel and Paras [5]	Discrete Bayesian filtering
1994	Mahler [22,23,34] *"Finite-Set Statistics"*	Random set theory
1996	Barlow et. al. [4] *"Unified Data Fusion"*	Heuristic
1996	Mahler-Kastella [13,16,40] *"Joint Multitarget Probabilities"*	"Finite-set statistics" under a new name and notation [24,29,30,40]
1997	Portenko et. al. [39]	Point processes

1.3 Summary of the Paper. The paper is organized as follows. In Section 2 we begin with a brief review of the basic elements of recursive Bayesian nonlinear filtering and estimation. In Section 3 we show that, when the number n of targets is unknown, even the most basic of issues--precise definition of the concept of a multitarget state--leads to trouble. In Section 4 we clarify the key issues underlying a special case--when the number n of targets is assumed *known*--by contrasting it with the EAMLE filter. In the sections that follow, we focus exclusively on the key issues associated with the unknown-n problem. In Section 5 we contrast the *filtering* aspects of the unknown-n problem with an *ad hoc* scheme called "generalized EAMLE." In Section 6 we contrast the *estimation* aspects of the unknown-n problem with "JMP"--which, as just noted, is a renaming of a truncated version of FISST. Section 7 describes the specific problems associated with stable convergence of two such estimators to the true solution. Conclusions may be found in Section 8.

2. Bayesian Single-Sensor, Single-Target Filtering and Estimation: A Review

Suppose that our goal is to determine the kinematic state $\mathbf{x}$ of a single moving target on the basis of point observations $\mathbf{z}_1,\ldots,\mathbf{z}_\alpha$ collected by a single sensor. In the discrete-time Bayesian framework we assume the following information:

- *Measurement space* = precise description of all measurements $\mathbf{z}$ that can be collected by the sensor
- *State space* = precise description of all parameters $\mathbf{x}$ necessary to uniquely describe all target states
- *Likelihood function*: $f(\mathbf{z} \mid \mathbf{x})$ = likelihood of seeing observation $\mathbf{z}$ given that the target has state $\mathbf{x}$
- Markov *motion model*: $f_{\alpha+1\mid\alpha}(\mathbf{x}_{\alpha+1} \mid \mathbf{x}_\alpha)$ = likelihood of target being in state $\mathbf{x}_{\alpha+1}$ at time-step $\alpha+1$, given that it was in state $\mathbf{x}_\alpha$ at time-step α
- *Initial state model*: $f_0(\mathbf{x}_0)$ = likelihood that target was initially in state $\mathbf{x}_0$
- *Posterior density*: $f_{\alpha\mid\alpha}(\mathbf{x}_\alpha \mid Z^\alpha)$ = likelihood that the target has state $\mathbf{x}_\alpha$ at time-step α, given that observations $Z_\alpha = \{\mathbf{z}_1,\ldots,\mathbf{z}_\alpha\}$ have been collected

2.1 Recursive Bayesian Nonlinear Filtering Equations. Given this and certain additional independence assumptions, the discrete-time Bayesian nonlinear filtering equations (7a) and (7b) show how to recursively propagate the posterior density as each measurement is collected [1,2,7,9,41].

2.2 State Estimators. The posterior density $f_{\alpha\mid\alpha}(\mathbf{x}_\alpha \mid Z^\alpha)$ contains all of the information that we need to estimate the state of the target at time-step α, but this information is unavailable for practical application unless we also have

a "mathematical can opener"--a *state estimator*--that enables us to extract it from the posterior. A (static) estimator of the state $\mathbf{x}$ is any family $\hat{\mathbf{x}}(\mathbf{z}_1,\ldots,\mathbf{z}_m)$ of state-valued functions of the (static) measurements $\mathbf{z}_1,\ldots,\mathbf{z}_m$. "Good" Bayes state estimators $\hat{\mathbf{x}}$ should be *Bayes-optimal* in the sense that, in comparison to all other possible estimators, they minimize the Bayes risk **[43, pp. 54-63]**

$$R_C(\hat{\mathbf{x}},m) = \int \cdots \int \int C(\mathbf{x}, \hat{\mathbf{x}}(\mathbf{z}_1,\ldots,\mathbf{z}_m))\, f(\mathbf{z}_1,\ldots,\mathbf{z}_m \mid \mathbf{x})\, f(\mathbf{x})\, d\mathbf{x}\, d\mathbf{z}_1 \cdots d\mathbf{x}_m$$

for some specified cost (i.e., objective or loss) function $C(\mathbf{x},\mathbf{y})$ defined on states $\mathbf{x},\mathbf{y}$. Secondly, they should be *statistically consistent* in the sense that $\hat{\mathbf{x}}(\mathbf{z}_1,\ldots,\mathbf{z}_m)$ converges to the actual target state $\mathbf{x}$ as the number of measurements m tends to infinity. Other properties (e.g., asymptotically unbiased, rapidly convergent, stably convergent, etc.) are desirable as well. In Bayesian theory, the two most common "good" estimators are the maximum *a posteriori* (MAP) and expected *a posteriori* (EAP) estimators:

$$\hat{\mathbf{x}}_{MAP}(\mathbf{z}_1,\ldots,\mathbf{z}_m) = \arg\sup_{\mathbf{x}} f(\mathbf{x} \mid \mathbf{z}_1,\ldots,\mathbf{z}_m), \qquad \hat{\mathbf{x}}_{EAP}(\mathbf{z}_1,\ldots,\mathbf{z}_m) = \int \mathbf{x}\, f(\mathbf{x} \mid \mathbf{z}_1,\ldots,\mathbf{z}_m)\, d\mathbf{x}$$

(where "*argsup*" means find the value(s) $\hat{\mathbf{x}}$ of $\mathbf{x}$ which maximize $f(\mathbf{x} \mid \mathbf{z}_1,\ldots,\mathbf{z}_m)$). Though it plays an important role in theory, the *EAP* estimator often results in unstable solutions in applications in which the posterior has multiple significant modes (e.g. low-SNR scenarios in which insufficient data has been collected to ensure the existence of a single dominant mode).

2.3 Provable Convergence of Estimators. An important point that is often overlooked is that the "goodness" of an estimator crucially depends on certain seemingly esoteric mathematical concerns. For example, Wald's proof **[44]** that the MAP estimator is consistent requires the following assumptions:

- The space of all measurements $\mathbf{z}$ is a *topological space* satisfying certain properties;
- the space of all states $\mathbf{x}$ is a *metric space* satisfying certain properties; and
- $f(\mathbf{z} \mid \mathbf{x})$ is *measurable* in the variable $\mathbf{z}$ (with respect to the measurement-space topology) and *continuous* in the variable $\mathbf{x}$ (with respect to the state-space metric).

3. What is a Multitarget State?

In the unknown-n problem we encounter basic conceptual difficulties even before getting started. Multitarget posterior densities $f(x_1,\ldots,x_n \mid Z)$ cannot even be defined unless one has at hand a multitarget measurement model $f(z_1,\ldots,z_k \mid x_1,\ldots,x_n)$ that tells us the likelihood of seeing measurements $z_1,\ldots,z_k$ given the presence of targets with states $x_1,\ldots,x_n$. However, we cannot even *define* a likelihood function unless we first:

- precisely define the state and measurement spaces,
- define topologies on the state and measurement spaces,
- define random variables on the state and measurement spaces using these topologies, and
- define the multitarget likelihood function as a conditional distribution of these random variables

But how do we uniquely specify all of the states that the multitarget system can occupy? Multitarget states must look something like this:

$\varnothing$: no-target state
$\mathbf{x}_1$: the single-target states
$\mathbf{x}_1,\mathbf{x}_2$: the two-target states
$\mathbf{x}_1,\ldots,\mathbf{x}_n$: the n-target states

However, this *naïve specification of multitarget states* is incomplete. The symbol $\mathbf{x},\mathbf{x}$ signifies not that a single

target with state $\mathbf{x}$ is present twice, but rather that two completely different targets happen to occupy the same kinematic state $\mathbf{x}$. Strictly speaking, therefore, the state of an individual target is not fully specified (in a multitarget context) unless a unique identifier (an "I.D. tag") has been attached to it, e.g.: $(\mathbf{x},\tau)$ where $\mathbf{x}$ is the target's kinematic state and τ is its unique identifying tag. Thus the incompletely specified two-target state $\mathbf{x},\mathbf{x}$ should be replaced by the completely-specified two-target state $(\mathbf{x},\tau_1)$, $(\mathbf{x},\tau_2)$. Two-target states of the form $(\mathbf{x},\tau)$, $(\mathbf{y},\tau)$ with $\mathbf{x} \neq \mathbf{y}$ must be excluded as non-physical since the same target cannot occupy two different kinematic states simultaneously. Also, note that the two-target state $(\mathbf{x}_1,\tau_1)$, $(\mathbf{x}_2,\tau_2)$ and the two-target state $(\mathbf{x}_2,\tau_2)$, $(\mathbf{x}_1,\tau_1)$ are not distinct: They represent the same two-target state. What results is a rather complicated state space which is partially discrete and partially continuous.

Finite-set statistics (FISST) resolves these technical problems by representing multitarget states *geometrically*--that is, as finite sets whose elements are the states of the individual targets, endowing multitarget states and multitarget observations with suitable topologies/metrics, and then showing that the resulting multitarget state space is sufficiently well-behaved to ensure good filtering behavior **[6, pp. 194-198]**.

4. Heuristic versus True Multitarget Likelihood Functions

When the number of targets is known, it would appear that the problems summarized in Section 3 can be sidestepped because, in this case, multitarget states can be modeled as ordinary vectors. This belief is mistaken. Given that we have multitarget likelihood function $f(z_1,\ldots,z_k \mid x_1,\ldots,x_n)$, how can we *systematically construct* it from a knowledge of the individual sensors--rather than assuming that it is either a heuristic contrivance or a mathematical abstraction that comes out of nowhere, *deus ex machina*? These key issues underlying known-n multitarget filtering are best illustrated by examinining the event-averaged maximum likelihood (EAMLE) filter described by Kastella in 1993. **[14]**

4.1 The EAMLE Filter. Suppose that a known number n of unknown targets is observed in Poisson-distributed clutter by a Gaussian tracking radar. If $\mathbf{x}_1,\ldots,\mathbf{x}_n$ are the states of the targets, concatenate them into a single "system-level" state vector $\mathbf{X} = (\mathbf{x}_1,\ldots,\mathbf{x}_n)$. Likewise, if $\mathbf{z}_1,\ldots,\mathbf{z}_m$ are the observations collected by the sensor, then concatenate them into a single "system-level" observation vector $\mathbf{Z} = (\mathbf{z}_1,\ldots,\mathbf{z}_m)$. Kastella contructs a heuristic multitarget "average likelihood" function

$$f_{EAMLE}(\mathbf{Z} \mid \mathbf{X},n) = \Sigma_\sigma f(\mathbf{Z} \mid \mathbf{X},n,\sigma)\, p(n,\sigma)$$

where the summation is over all possible associations σ between states $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and measurements $\mathbf{z}_1,\ldots,\mathbf{z}_m$ and where $f(\mathbf{Z} \mid \mathbf{X},n,\sigma)$ is the likelihood given a particular association σ. Assume that the between-measurements evolution of the "system" state vector $\mathbf{X}$ can be modeled by a conventional Kalman equation $\mathbf{X}_{k+1} = A_k\mathbf{X}_k + \mathbf{V}_k$. If $f_0(\mathbf{X} \mid n)$ denotes prior knowledge of the system state vector $\mathbf{X}$, then--assuming that $f_{EAMLE}(\mathbf{Z} \mid \mathbf{X},n)$ is a true multitarget likelihood function--$argsup_{\mathbf{X}} f(\mathbf{Z} \mid \mathbf{X},n)\, f_0(\mathbf{X} \mid n)$ provides an optimal estimate of the system-level state $\mathbf{X}$. Using these measurement and motion models, the multitarget system vector $\mathbf{X}$ can be tracked by an EKF.

This discussion implies that known-n filtering can be accomplished entirely within a "naïve" vector-based paradigm based on standard filtering theory. In reality, such a perspective overlooks a key point:

- Is $f_{EAMLE}(\mathbf{Z} \mid \mathbf{X},n)$ a *true* multitarget likelihood function or is it an *ad hoc*, purely heuristic contrivance?

4.2 True Likelihood Functions: Single-Target Case. For purposes of illustration, consider a simpler problem: single-target filtering. Here, the likelihood function $f(\mathbf{z} \mid \mathbf{x})$ does not appear out of nowhere. We are able to *systematically construct* it from an explicit sensor model $\mathbf{z} = h(\mathbf{x},\mathbf{w})$ that describes the way a target with state $\mathbf{x}$ generates sensor measurements $\mathbf{z}$ under the influence of random noise $\mathbf{w}$ described by a probability density $f_{\mathbf{w}}(\mathbf{y})$. For example, let C be a matrix and suppose that the sensor model is $\mathbf{z} = C\mathbf{x} + \mathbf{w}$. Then

$$\int_S f_z(\mathrm{y} \mid \mathrm{x})\, dy \;=\; Pr(\mathrm{z} \in S) \;=\; Pr(C\mathrm{x}+\mathrm{w} \in S) \;=\; Pr(\mathrm{w} \in S\text{-}C\mathrm{x}) \;=\; \int_{S\text{-}C\mathrm{x}} f_{\mathrm{w}}(\mathrm{y})\, dy \;=\; \int_S f_{\mathrm{w}}(\mathrm{y}\text{-}C\mathrm{x})\, dy$$

where $S\text{-}C\mathrm{x}$ denotes the set of all $\mathrm{s} - C\mathrm{x}$ with $\mathrm{s} \in S$. Since this is true for all measurable S, the likelihood function is $f_z(\mathrm{y} \mid \mathrm{x}) = f_{\mathrm{w}}(\mathrm{y}\text{-}C\mathrm{x})$ almost everywhere in y. In other words: We know that $f_z(\mathrm{y} \mid \mathrm{x})$ is the *true likelihood function* for the problem (and not just a heuristic contrivance) because it can be *systematically and rigorously constructed from the actual sensor model using standard probabilistic reasoning*. (In purely mathematical language: The likelihood function $f_z(\mathrm{y} \mid \mathrm{x})$ can be explicitly constructed as the Radon-Nikodým derivative $f_z = dp_z/d\lambda$, with respect to Lebesgue measure λ, of probability measure $p_z(S) = Pr(\mathrm{z} \in S)$.) In like fashion we ask:

- How do we construct *multitarget sensor models* of the general form $Z = h(X,W)$ where Z is a multitarget measurement, X is a multitarget state, and W is a random clutter process?
- How do we rigorously, systematically construct multitarget likelihood functions $f(Z \mid X)$ from the model $Z = h(X,W)$ using standard probabilistic procedures?

4.3 FISST Multitarget Likelihood Functions. Addressing these questions is one of the advances made possible by FISST. The basic idea is this: We construct an explicit *multitarget measurement model* of the form

$$Z = T \cup W$$

where the randomly varying finite set $T = T(X)$ models the observations generated by the actual targets and where the randomly varying finite set W models the false alarm and/or clutter observations. Assume that spurious observations are independent of actual observations. Then

$$\int_S f_Z(Y \mid X)\, \delta Y \;=\; Pr(Z \subseteq S) \;=\; Pr(T \cup W \subseteq S) \;=\; \int_S \Sigma_{E \subseteq Y} f_T(Y\text{-}E \mid X)\, f_W(E)\, \delta Y$$

where the integrals are multi-object "set integrals" (see **[6, pp. 141-144; 34, p. 334]** and equation (15) below). Since this is true for all measurable S, we get

$$f_Z(Y \mid X) \;=\; \Sigma_{E \subseteq Y} f_T(Y\text{-}E \mid X)\, f_W(E)$$

where $f_T(Y \mid X)$ can be explicitly constructed from the underlying sensor measurement model $\mathrm{z} = h(\mathrm{x},\mathrm{w})$, and where $f_W(E)$ can be explicitly constructed from an underlying clutter measurement model $\mathrm{z} = g(\mathrm{x},\mathrm{v})$. Indeed, as early as November 1993, Mahler showed how to use the "set derivative," **[6, pp. 150-152,20,22,27,34]** to systematically and rigorously construct multitarget likelihood functions $f_Z(Z \mid X)$, $f_T(Z \mid X)$, $f_W(Z)$ from underlying sensor models *even when the number of targets is unknown*. Consequently, just as the equation $Pr(\mathrm{z} \in S) = \int_S f(\mathrm{y} \mid \mathrm{x})\, dy$ is the fundamental relationship that allows us to conceptually link single-target sensor measurement models with likelihood functions, so the FISST equation $Pr(Z \subseteq S) = \int_S f(Y \mid X)\, \delta Y$ is the fundamental relationship that allows us to conceptually link multitarget sensor measurement models with "global" (i.e., true multitarget) likelihood functions. In particular: In April 1994 we demonstrated **[6, pp. 228-235; 20]** that, when FISST is restricted to the special situation addressed by EAMLE then

(8) $$f_{FISST}(\mathrm{z}_1, \ldots, \mathrm{z}_m \mid \mathrm{x}_1, \ldots, \mathrm{x}_n) \;=\; m!\, f_{EAMLE}(\mathrm{z}_1, \ldots, \mathrm{z}_m \mid \mathrm{x}_1, \ldots, \mathrm{x}_n, n)$$

That is: Using FISST, one can prove that the heuristic multitarget likelihood $f_{EAMLE}(\mathrm{Z} \mid \mathrm{X}, n)$ actually is the true multitarget likelihood function for this particular multitarget problem.

5. Statistically Coupled versus Statistically Decoupled Multitarget Probability Distributions

We now turn exclusively to applications in which the number n of targets is *not* known. Some of the key issues involved in the unknown-n problem are best illustrated by describing "generalized EAMLE", a multitarget filtering

scheme devised at Lockheed Martin by Kastella in May 1993 [17] that fails to exhibit any grasp of these issues.

5.1 "Generalized EAMLE" Filter. Generalized EAMLE consists of the following *ad hoc* sequence of Bayesian and non-Bayesian steps:

Step 1: Let $f_0(\mathbf{X} \mid n)$ be the prior over all system state vectors $\mathbf{X}$ with n targets, i.e. $\int f_0(\mathbf{X} \mid n)\, d\mathbf{X} = 1$
Step 2: Let $f(\mathbf{Z} \mid \mathbf{X}, n)$ be the EAMLE average-likelihood function ($\mathbf{X}$ is a system state vector with n targets)
Step 3: Construct a mean-likelihood function $f(\mathbf{Z} \mid n) = \int f(\mathbf{Z} \mid \mathbf{X}, n)\, f_0(\mathbf{X} \mid n)\, d\mathbf{X}$ for each $n = 1,2,\ldots$
Step 4: Determine the most likely value of n via an ML estimator: i.e. $\hat{n} = argmax_n\, f(\mathbf{Z} \mid n)$
Step 5: Given $\hat{n}$, use a MAP estimator to find the value of the multitarget state $\mathbf{X}$, i.e.,

$$\hat{\mathbf{X}}\Big|_{n=\hat{n}} = \underset{\mathbf{X}:n=\hat{n}}{argsup}\, f(\mathbf{Z} \mid \mathbf{X}, \hat{n})\, f_0(\mathbf{X} \mid \hat{n})$$

(Since the only description of this approach is an unpublished technical document, we reproduce the relevant text as an Appendix.) Generalized EAMLE is "horizontal" in structure: The unknown-n multitarget estimation problem is dissected into parallel, statistically decoupled known-n problems for $n = 1,2,\ldots,$ with each known-n problem being processed using one cycle of an EAMLE filter (see Figure 1). Multitarget state space is partitioned into separate state spaces, one for each target number $n = 1,2,\ldots$ (the zero-target case is overlooked entirely). The likelihood functions $f(\mathbf{Z} \mid \mathbf{X}, n)$ for $n = 1,2,\ldots$ presume that the number of targets is fixed. Prior knowledge is specified by separate prior probability densities $f_0(\mathbf{X} \mid n)$ for $n = 1,2,\ldots$ Target number is determined not by directly comparing all multitarget states $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ with all multitarget states $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_{n'})$ for arbitrary target number n, n' but, rather, by computing mean-likelihoods $f(\mathbf{Z} \mid n)$ separately for each known-n problem and then in effect comparing an "typical n-target state" to a "typical n'-target state". This statistical decoupling of the unknown-n problem has the following consequences:

- Because the priors $f_0(\mathbf{X} \mid n)$ are decoupled, they contain no prior knowledge $f_0(n)$ regarding the number of targets
- Because the parallel known-n filters are decoupled, probability mass cannot shift from one row of Figure 1 to another.
- Because state estimation is decoupled, multitarget states with differing target number cannot be directly compared with each other to more effectively determine which are most or least likely.

5.2 Key Points of Multitarget Recursive Bayesian Filtering. The source of these difficulties is the fact that generalized EAMLE exhibits no conceptual grasp of the following key points, especially item (11):

(9) Reconceptualize a group of unknown targets, of unknown number, as a *single joint multitarget system*
(10) Reconceptualize target number and individual target states as just specific parameters of a *single joint multitarget state*
(11) Reconceptualize priors, posteriors, and likelihoods as *single densities involving an unknown number of targets and target states*
(12) Reconceptualize detection and estimation as different aspects of a *single, simultaneous statistical process*
(13) Recognize that one cannot just "declare victory": reconceptualization is the easy part, because it leads to unexpected theoretical and practical difficulties that cannot be swept under the rug

5.3 FISST Statistically Coupled Filtering. These key points are the conceptual foundation of true (i.e., non-naïve) multitarget Bayesian recursive nonlinear filtering--and the FISST approach to it in particular. This approach is vertical in structure: It correctly models the statistical coupling between all multitarget states and--in particular--regards multitarget priors and posteriors as *single probability distributions* (see Figure 2). For example, the FISST "global maximum likelihood estimator (global MLE)" [20,22,27,34] was introduced in Nov. 1994. The global

$$\begin{array}{llllll}
n=1 & f_0(\mathrm{X}\mid 1) & \rightarrow & f(\mathrm{Z}\mid \mathrm{X},1)\, f_0(\mathrm{X}\mid 1) & \rightarrow & f(\mathrm{Z}\mid 1) \\
n=2: & f_0(\mathrm{X}\mid 2) & \rightarrow & f(\mathrm{Z}\mid \mathrm{X},2)\, f_0(\mathrm{X}\mid 2) & \rightarrow & f(\mathrm{Z}\mid 2) \\
 & & & \cdots & & \\
n=k: & f_0(\mathrm{X}\mid k) & \rightarrow & f(\mathrm{Z}\mid \mathrm{X},k)\, f_0(\mathrm{X}\mid k) & \rightarrow & f(\mathrm{Z}\mid k)
\end{array}$$

Figure 1: "Generalized EAMLE" The conceptual key points underlying multitarget Bayesian nonlinear filtering--unified multitarget state spaces, joint multitarget probability distributions, and simultaneous estimation of multitarget states--are illustrated via comparison with "generalized EAMLE," an *ad hoc* approach that fails to grasp these points. Generalized EAMLE has a "horizontal" structure (above): The multitarget filtering problem is broken up into parallel, completely decoupled filters. Each filter deals with a different hypothesis about the number $n = 1,\ldots,k$ of targets (the zero-target hypothesis is neglected). Because the priors $f_0(\mathbf{X} \mid n)$ are decoupled, they provide no means of specifying initial belief $f_0(n)$ regarding the number of targets. Because the parallel filters are decoupled, probability mass cannot shift from one row to another. Because state estimation is decoupled (i.e., target number is determined by computing a marginal likelihood $f(\mathbf{Z} \mid n)$ for each row), multitarget states with differing target number cannot be directly compared with each other to determine which are most or least likely.

$$\begin{bmatrix} f_0(0) \\ f_0(\mathrm{X}\mid 1) f_0(1) \\ f_0(\mathrm{X}\mid 2) f_0(2) \\ \cdots \\ f_0(\mathrm{X}\mid k) f_0(k) \end{bmatrix} \rightarrow \begin{bmatrix} f(\mathrm{Z}\mid 0) \\ f(\mathrm{Z}\mid \mathrm{X},1) \\ f(\mathrm{Z}\mid \mathrm{X},2) \\ \cdots \\ f(\mathrm{Z}\mid \mathrm{X},k) \end{bmatrix} \cdot \begin{bmatrix} f_0(0) \\ f_0(\mathrm{X}\mid 1) f_0(1) \\ f_0(\mathrm{X}\mid 2) f_0(2) \\ \cdots \\ f_0(\mathrm{X}\mid k) f_0(k) \end{bmatrix} \rightarrow \begin{bmatrix} f(0\mid \mathrm{Z}) \\ f(\mathrm{X}\mid \mathrm{Z},1) f(1\mid \mathrm{Z}) \\ f(\mathrm{X}\mid \mathrm{Z},2) f(2\mid \mathrm{Z}) \\ \cdots \\ f(\mathrm{X}\mid \mathrm{Z},k) f(k\mid \mathrm{Z}) \end{bmatrix}$$

$$f_0(X) \quad \rightarrow \quad f(\mathrm{Z}\mid \mathrm{X}) \cdot f_0(X) \quad \rightarrow \quad f(X\mid \mathrm{Z})$$

Figure 2: Multitarget Bayesian nonlinear filtering, by way of contrast, has a "vertical" structure (above) that correctly models the coupling between (i.e., relative probability contributions of) all multitarget states. It does so via multitarget densities that are "global" ("joint") in that they are *single densities involving an unknown number of targets and target locations*, defined over a *unified multitarget state space*. The multitarget prior $f_0(X)$ is a "global" collection $\{f_0(\mathbf{X} \mid i) f_0(i)\}_{i=0,1,\ldots,k}$ of conventional densities coupled together by a joint normality condition $\int f_0(X)\, \delta X = 1$ (where ' $\int$ ' denotes a multitarget or "set" integral). The multitarget posterior $f_0(X \mid Z)$ is a "joint multitarget probability" $\{f_0(\mathbf{X} \mid \mathbf{Z}, i)\ f(i \mid \mathbf{Z})\}_{i=0,1,\ldots,k}$ of the same type. The joint multitarget likelihood function $f(Z \mid X)$ is a joint collection $\{f(\mathbf{Z} \mid \mathbf{X}, i)\}_{i=0,1,\ldots,k}$.

MLE integrated the functions of detection and localization into a single, simultaneous, and *provably optimal* statistical operation. Suppose the sensor collects observation-sets $Z^{(m)}: Z_1, \ldots, Z_m$. Then the *global likelihood function* is $L(X \mid Z^{(m)}) = f_\Sigma(Z_1 \mid X) \cdots f_\Sigma(Z_m \mid X)$--i.e., a single density involving an unknown number of targets and target states. The global MLE is **[22]**:

$$\{\hat{x}_1, \ldots, \hat{x}_{\hat{k}}\} \qquad \underset{X}{argsup}\ L(X \mid Z^{(m)}) \;=\; \underset{n, x_1, \ldots, x_n}{argsup}\ L(\{x_1, \ldots, x_n \quad Z^{(m)})$$

where "$argsup_X$" denotes that unique value of X which (if it exists) maximizes $L(X \mid Z^{(m)})$. It is provably optimal because $f(Z \mid X)$ is a true multitarget likelihood function (since it can be computed *directly from sensor characteristics* using the "set derivative" **[6,22,34]**). In early 1994, three Lockheed Martin internal technical reports described the global MLE process using simple numerical examples. **[6, pp. 256-259; 27]** In April 1994, EAMLE was proved to be a special case of the global MLE. **[20]** Also in November 1993, Mahler introduced the concept of a "global" probability density, i.e. one which takes account of the fact that both numbers of targets and numbers of observations are generally random. Using both set notation and more conventional notation, it was shown that such densities could be equivalently written in the form

$$(14) \qquad f(\{x_1, \ldots, x_n\}) = n!\, f(x_1, \ldots, x_n)$$

and that, in either notation, they were defined by the joint normality condition **[20 p. 9]**

$$(15) \qquad \sum_{n=0}^{k} \int f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \;=\; 1 \;=\; \sum_{n=0}^{k} \frac{1}{k!} f(\{x_1, \ldots, x_n\})\, dx_1 \cdots dx_n$$

that was subsequently expressed in abbreviated form as a so-called "set integral": $\int f(X \mid Z^{(m)})\, \delta X = 1$.

6. Optimal versus Heuristic Multitarget Estimators

Thus far we have underscored difficulties associated with the filtering aspects of equations (7a′, 7b′). Even if we have managed to construct the multitarget posterior, however, our work is still not done. We must also find a means of extracting from it the information that we want: estimates of the multitarget state variables (target number, identities, positions, velocities, etc.). Here, we encounter even greater difficulties. Some of the key points involved in unknown-*n* multitarget state estimation are best illustrated by contrasting them to "joint multitarget probabilities (JMP)", a renaming of FISST core concepts that fails to address these key points.

6.1 "Joint Multitarget Probabilities (JMP)". Describing it as "similar in some respects to the random set formalism of Mahler" **[13 p. 168]**, Kastella introduced "JMP" **[13,16]** at Lockheed Martin in July 1996 as a solution to the unknown-*n* problem. However, only a trivial change of notation differentiates "JMP" from a number of core FISST concepts devised years earlier. **[24 Sect's 2-3,29,30,40 pp. 27-28]** The relevant chronology is listed in Table III below. From late 1993 on, research on FISST was widely reported in the form of simple numerical examples and other internal Lockheed Martin research reports; internal and external presentations; and published conference papers. When it appeared in July 1996, "JMP" (as a solution to both unknown-*n* multitarget filtering and multitarget information theory) reiterated FISST "key points" developed during this period (indeed, even some of the same notation--see equ. (14) above). These include:

- the concept of "a single density involving an unknown number of targets and target locations"
- multitarget posteriors
- set integrals
- multitarget Kullback-Leibler measures

- the "almost-parallel worlds principle (APWOP)," a systematic approach to generalizing single-sensor, single-target concepts and approaches to multisensor-multitarget problems

In particular, Kastella used the APWOP to directly generalize his single-target "discrimination gain" approach to the multitarget case. **[13, p. 168,24,40 pp. 27-28]**

TABLE III: A FISST Chronology

Date	Finite-Set Statistics	"Joint Multitarget Probabilities"	
05/93		Generalized EAMLE [17] ←	
11/93	Global MLE **[20,22]**		
03/94	Numerical examples **[27]**		
04/94	EAMLE = special case of global MLE **[20]**		
04/94	Joint multi-object densities **[20]**		2½ yrs
10/94	Multitarget posteriors, multitarget Kullback-Leibler metrics, APWOP **[34]**		
11/95		Unknown-n unsolved **[15]** ←	
03/96	Multitarget nonlinear filtering **[23]**		
07/96		"JMP" **[13]**	

Given this chronology and our discussion in Section 4, any claim that "JMP" is just an obvious embellishment of "generalized EAMLE" (and therefore that "JMP" actually pre-dates FISST) would simply not be credible. On the one hand, in Section 4 we demonstrated that generalized EAMLE is an *ad hoc* procedure that exhibits no understanding of any of the conceptual "key points" (9)-(13) of multitarget Bayesian recursive nonlinear filtering (and of "JMP" in particular). On the other hand, as late as November 1995, Kastella reported being still in search of a satisfactory solution of the unknown-n problem--despite the prior existence of "generalized EAMLE". (To wit: "Some basic questions for future research are...How to extend the results to the case of an unknown number of targets..." **[15 p. 43]**). It would be no more credible to claim that "JMP" is "simpler" than FISST. A trivial change of notation constitutes neither a simplification nor a technical advance. Neither does inattentiveness to unexpected research difficulties--for example, recognition that optimal estimation in multitarget Bayesian filtering cannot be taken for granted; or recognition of the closely related fact that continuous variables create fundamental practical difficulties in multitarget Bayesian filtering. **[6 pp. 190-194,23,29]**

Specifically, in "JMP" state space is completely discretized, with the following assertion that the continuous-state case follows immediately from a trivial change of notation:

- "For simplicity, the target space is discretized into a collection of cells (in the continuous case, the cell probabilities can be replaced by densities in the usual way)." **[16 pp. 123]**

(In actuality, passage from the discrete-variable to the continuous-variable case introduces fundamental complications. **[6 pp. 190-194;18]**) Given this, a JMP has the form **[13 p. 167,16 p. 122]**

$$(16) \qquad p(c_1,\ldots,c_n \mid Z) \qquad \text{or} \qquad p^n_{c_1,\ldots,c_n}(x_1,\ldots,x_n \mid Z)$$

where Z is a set of observations, n is the (unknown) number of targets, $x_1,\ldots,x_n$ are discrete location cells, and $c_1,\ldots,c_n$ are the respective I.D. classes of the targets in these cells. This is essentially the alternative notation for a global posterior density described in equations (8) and (14). In JMP, the family of probability distributions p^n indexed by n are treated as a single joint probabilistic entity by requiring **[13 p. 168,16 p. 123]** that they satisfy the following normalization criterion:

$$(17) \qquad \sum_{n=0}^{\infty} \sum_{\substack{c_1,\ldots,c_n \\ x_1,\ldots,x_n}} p^n_{c_1,\ldots,c_n}(x_1,\ldots,x_n \mid Z) \quad = \quad 1$$

This is the earlier FISST formula (15) assuming that all state variables are discrete. That is, a JMP is a FISST global posterior density written in terms of the following alternative notation (see equ.'s (8,14)):

$$(18) \quad p(c_1,\ldots,c_n \mid Z) \;=\; \frac{1}{n!} f(\{c_1,\ldots,c_n\} \mid Z)\,, \quad p^n_{c_1,\ldots,c_n}(x_1,\ldots,x_n \mid Z) \;=\; \frac{1}{n!} f(\{(x_1,c_1),\ldots,(x_n,c_n)\} \mid Z)$$

"JMP" employs the multitarget filtering equations (7a′, 7b′) with a simplified multitarget motion model (5′):

$$(19) \qquad f^n_{\alpha+1|\alpha}(\mathbf{y}_1,\ldots,\mathbf{y}_n \mid \mathbf{x}_1,\ldots,\mathbf{x}_n) \;=\; f_{\alpha+1|\alpha}(\mathbf{y}_1 \mid \mathbf{x}_1) \cdots f_{\alpha+1|\alpha}(\mathbf{y}_n \mid \mathbf{x}_n)$$

This motion model assumes that the motions of the targets in a scenario are independent and that targets do not enter or leave the scene. (In the second prequel to this paper **[30]** we pointed out that dynamic multitarget scenarios, in which targets can leave or enter the scene, require general multitarget motion models that more completely capture the actual complexity of real tracking problems.) Target class and kinematics variables are assumed to be independent. The multitarget state is computed using the following multitarget estimator:

$$(20) \quad \{\hat{x}_1,\ldots\hat{x}_{\hat{n}}\} \;=\; \underset{x_1,\ldots,x_{\hat{n}}}{argsup}\; p^{\hat{n}}(x_1,\ldots,x_{\hat{n}} \mid Z) \qquad \text{where} \qquad \hat{n} \;=\; \underset{n\geq 0}{argmax} \int p^n(x_1,\ldots,x_n \mid Z)\, dx_1,\ldots,dx_n$$

6.2 Key Points of Multitarget State Estimation. Once again, this discussion implies that unknown-*n* filtering can be accomplished entirely within a purely heuristic generalization of standard filtering theory. In reality, a naïve approach overlooks several key points:

- Is the proposed estimator equ. (20) optimal or is it an *ad hoc*, heuristic contrivance?
- In particular: Is it Bayes-optimal and is it guaranteed to converge to the correct solution?
- Computation of the marginal (Step 1) throws away information contained in the multitarget posterior. What are the consequences of this for rapidity of convergence?
- What are the stability properties of this estimator under high-noise or high-clutter conditions?

In mid-1995 Mahler proved that the multitarget estimator equ. (20) is Bayes-optimal (the proof is reported in *Mathematics of Data Fusion* **[6 pp. 190-205]**.) However, it is unclear whether or not this estimator is convergent. In Section 7 below, we demonstrate that it is likely to be slowly convergent and to produce unstable solutions under low-observable conditions. At that time, he also proved that another estimator is both provably Bayes-optimal and provably convergent (see Section 7).

To more fully understand these key points, let us take the position that a naïve, heuristic approach is sufficient and, consequently, that we can *proceed immediately to a declaration of victory*. We assume that we can define a multitarget posterior distribution function on a multitarget state space that need not be carefully defined. To keep things simple, assume that all targets are motionless, exist in one dimension, and are completely specified by their locations on the real line as measured in *meters*. Also assume that a single sensor collects a set Z of measurements from the targets, whose number as well as positions are unknown and are to be estimated. One can write a *naïve posterior distribution function f(state* | *Z)* on the multitarget state space, given measurements Z, as follows:

$$
\begin{aligned}
f(\emptyset \quad Z) &= \text{posterior likelihood of zero targets} \\
f(x_1 \mid Z) &= \text{posterior likelihood of one target in state } x_1 \\
f(x_1, x_2 \mid Z) &= \text{posterior likelihood of two targets in states } x_1, x_2 \\
f(x_1, \ldots, x_n \mid Z) &= \text{posterior likelihood of } n \text{ targets in states } x_1, \ldots, x_n
\end{aligned}
$$

Since the cumulative likelihood summed over all multitarget states must be 1, it follows that

$$f(0 \mid Z) + f(1 \mid Z) + \cdots + f(n \mid Z) + \cdots + f(M \mid Z) = 1$$

where $f(0 \mid Z) = f(\emptyset \mid Z)$ and $f(n \mid Z) = \int f(x_1, \ldots, x_n \mid Z)\, dx_1 \cdots dx_n$ for $n = 1, \ldots, M$, and where M is the maximum expected number of targets in the scenario. Now, let us use a *naïve MAP procedure* to estimate the complete state of the multitarget system. Then we would write:

$$\{\hat{x}_1, \ldots, \hat{x}_{\hat{n}}\} \qquad \underset{n,\, x_1, \ldots, x_n}{\mathit{argsup}}\; f(x_1, \ldots, x_n \mid Z)$$

where $\hat{n}$ is the estimated number of targets and the $\hat{x}_i$ are the estimated positions of the $\hat{n}$ targets. Unfortunately, there is a problem: From the definition of a Riemann integral we know that:

- $f(0 \mid Z)$ = unitless probability
- $f(1 \mid Z)$ = unitless probability, so units of $f(x_1 \mid Z)$ must be $1/meter$ since the units of dx_1 are meters
- $f(2 \mid Z)$ = unitless probability, so units of $f(x_1, x_2 \mid Z)$ must be $1/meter^2$
- $f(n \mid Z)$ = unitless probability, so units of $f(x_1, \ldots, x_n \mid Z)$ must be $1/meter^n$

Consequently the "*argsup*" operation is not defined since the quantities $f(\emptyset \mid Z), f(x_1 \mid Z), \ldots, f(x_1, \ldots, x_n \mid Z), \ldots$ are *incommensurable with respect to units of measurement*. Thus the MAP estimator *cannot even be defined!*

One might try to sidestep this problem by using *Riemann-Stieltjes* integrals. That is, let $g(x)$ be an arbitrary density function on (single-target) state space and let $G(x) = \int_0^x g(y)\, dy$ be its corresponding cumulative probability function. Then one could instead define different multitarget posterior densities $h(x_1, \ldots, x_n \mid Z)$ using Riemann-Stieltjes integrals

$$h(n \mid Z) \qquad \int h(x_1, \ldots, x_n \mid Z)\, dG(x_1) \cdots dG(x_n) \;=\; \int h(x_1, \ldots, x_n \mid Z)\, g(x_1) \cdots g(x_n)\, dx_1 \cdots dx_n$$

where, now, the multitarget distributions

$$h(x_1, \ldots, x_n \mid Z) = \frac{f(x_1, \ldots, x_n \mid Z)}{g(x_1) \cdots g(x_n)}$$

are unitless. Then, the multitarget distribution h will not have the incommensurability-of-units problem just noted and so could be used to define a multitarget MAP estimate. The price, however, is the unacceptable introduction of an *arbitrary "fudge factor"*--the density g--into the concept of a multitarget posterior distribution h.

If we instead turn to the EAP estimator for our salvation, our troubles get even worse. A multitarget posterior expectation, if it exists, must have the general form $\int \langle x_1, \ldots, x_n \rangle f(\langle x_1, \ldots, x_n \rangle \mid Z)\, d\langle x_1, \ldots, x_n \rangle$ where the (as yet to be defined) integral is taken over all (thus far vaguely defined) multitarget states $\langle x_1, \ldots, x_n \rangle$. Such an integral cannot even be defined unless, at minimum, the multitarget state space is a vector space--in particular, unless it has a concept of addition/subtraction. But how does one add the zero-target state $\emptyset$ to a single-target state x? Or

a single-target state $\mathbf{x}$ to a two-target state $\mathbf{x}_1, \mathbf{x}_2$? We could attempt to address this problem by embedding the multitarget state space in a larger, enveloping space which *is* a vector space (as indeed can be done in many ways). In this case, however, there is no guarantee that the posterior expectation would yield values which are actual multitarget states. Rather, it is more likely that it would yield values which are contained in the enveloping vector space and therefore which would have no physical meaning. In summary:

- *Having been denied the accustomed security of the classical estimators, we are forced to propose new ones and demonstrate that they are statistically well-behaved.*

Even if a multitarget MAP or EAP estimator could be defined, to prove convergence of the multitarget filter we would still need to know that the multitarget likelihood function is *measurable* in the variables $z_1, \ldots, z_k$ and *continuous* in the variables $x_1, \ldots, x_n$. But measurable with respect to *which topology* on the multitarget measurement space? Continuous with respect to *which metric* on the multitarget state space? The latter question is far from trivial. What is the distance between a single-target state x and a two-target state x_1, x_2? Or the zero-target state $\varnothing$ and the single-target state x_1? What is the distance between the two-target states x_1, x_2 and x_2, x_1 if $x_1 \neq x_2$? (The Euclidean metric gives us $\| (x_1, x_2) - (x_2, x_1) \| = \| (x_1 - x_2, x_2 - x_1) \| \neq 0$. But the state "an F-16 at x_1 flown by Joe and an F-22 at x_2 flown by Ralph" is the same multitarget state as "an F-22 at x_2 flown by Ralph and an F-16 at x_1 flown by Joe".) One can then try to tinker various metrics for multitarget state space, only to get pulled into a a morass of arbitrary, *ad hoc* definitionizing.

6.3 FISST Optimal Multitarget Estimation. Resolution of such questions is one of the advances made possible by FISST. The unexpected research difficulties associated with continuous- and discrete-variable "global" (i.e., joint multitarget) posterior and prior distributions were quickly recognized following their introduction in Dec. 1994 **[34]**. These difficulties were resolved, by July 1995, via the definition of two new *MAP-like multitarget estimators* and verification of their respective optimality properties. These results were reported in *Mathematics of Data Fusion* **[6 pp. 194-205]**. These estimators are:

- *Marginal Multitarget (MaM) State Estimator* (Provably Bayes-optimal; unproved convergence)

$$\{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{\hat{n}}\} \qquad \underset{\mathbf{x}_1, \ldots, \mathbf{x}_{\hat{n}}}{argsup}\, f(\{\mathbf{x}_1, \ldots, \mathbf{x}_{\hat{n}} | Z) \qquad \text{where} \qquad \hat{n} \qquad \underset{n}{argmax}\, \frac{1}{n!} \int f(\mathbf{x}_1, \ldots, \mathbf{x}_n | Z)\, d\mathbf{x}_1 \cdots d\mathbf{x}_n$$

- *Joint Multitarget (JoM) State Estimator* (Provably Bayes-optimal and convergent): for some constant c,

$$\{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{\hat{n}}\} \qquad \underset{n, \mathbf{x}_1, \ldots, \mathbf{x}_n}{argsup}\, f(\{\mathbf{x}_1, \ldots, \mathbf{x}_n | Z) \cdot c^n$$

(Note: In *Mathematics of Data Fusion*, *MaM* was called "GMAP-I" and *JoM* was called "GMAP-II".

7. Stable Convergence of Multitarget Estimators

The purpose of this section is to describe some of the difficulties associated with multitarget estimation and to assess the relative merits of the *JoM* and *MaM* estimators described in Section 6.3. We will begin with an intuitive discussion in Section 7.1 that illustrates why the *MaM* estimator can be expected to have less desirable performance than the *JoM* estimator. Then, in Section 7.2, we will verify this assessment by analytically comparing the convergence properties of these two estimators in a simple model problem.

7.1 Why Estimation Using Marginal Distributions Leads to Trouble. Under low-observable conditions special care must be exercised in the selection of state estimators in both single-target and multitarget problems. Consider the following analogy. The following joint posterior

$$f(x,y) = \tfrac{1}{2} N_{1.0}(x)\, N_{0.1}(y-2) + \tfrac{1}{2} N_{0.001}(x-5)\, N_{2.0}(y)$$

describes the current position of a target on the x-y plane, where $N_\sigma(x)$ denotes a one-dimensional normal distribution with variance σ^2. This distribution is bimodal with a large dominant mode near $(5,0)$ and a small minor mode near $(0,2)$. The true *joint* estimate of (x,y) is the joint MAP estimate

$$(\hat{x}_{MAP}, \hat{y}_{MAP}) = argsup_{x,y} f(x,y) = (5,0)$$

The EAP estimate is $(\hat{x}_{EAP}, \hat{y}_{EAP}) = (0,2) + (5,0) = (5,2)$. This estimate, which lies on a low-probability part of the tail of the dominant mode, is not very good. The reason is that the value of the EAP estimator is strongly influenced by the small minor mode. Suppose, instead, that we compute the marginal posteriors

$$f(y) = \quad f(x,y)\, dx = \tfrac{1}{2} N_{0.1}(y-2) + \tfrac{1}{2} N_{2.0}(y), \qquad f(x) = \int f(x,y)\, dy = \tfrac{1}{2} N_{1.0}(x) + \tfrac{1}{2} N_{0.001}(x-5)$$

Notice that x,y are correlated since $f(x,y) \neq f(x)\, f(y)$. Three possible marginal estimates can be derived:

$$(argsup_x f(x),\ argsup_y f(y)) = (5,\ argsup_y f(5,y)) = (argsup_x f(x,2),\ 2) = (5,2)$$

As with the EAP estimate, these estimates are strongly influenced by the small minor mode.

Now, under low-observable conditions a posterior $f(x,y \mid Z)$ will be highly *multi-modal.* Even if it has a single well-localized dominant mode, there will be a number of small minor "clutter" modes that will tend to jump erratically from time-step to time-step. If we use the EAP estimator or any of the above marginal-based estimators, the previous example shows that their values will all tend to unstable (i.e., vary erratically) whereas the joint MAP estimate will follow the dominant mode and therefore tend to be coherent. The same reasoning applies even if one of the state variables is discrete--for example, if the posterior $f(x,\omega \mid Z)$ where x is position (continuous) and ω is target identity (discrete). If x and ω are correlated $(f(x,\omega) \neq f(x)\, f(\omega))$ then deducing kinematics and target I.D. from the marginals $f(\omega) = \int f(x,\omega)\, dx$, $f(x) = \Sigma_\omega f(x,\omega)$ will lead to the same kinds of instability.

In fact, the same reasoning applies to determination of target number and target kinematics/I.D. in the multitarget case. If target number is correlated with target states, basing estimates on the marginal posterior $f(0 \mid Z) = f(\emptyset \mid Z)$, $f(n \mid Z) = \int f(x_1,\ldots,x_n \mid Z)\, dx_1 \cdots dx_n$ will lead to instability under low-observable conditions. The following example highlights this behavior. Suppose that we are given the multitarget posterior:

$$f(\emptyset \mid Z) = 1-q, \qquad f(x \mid Z) = q\, N_\sigma(x)$$

where σ is very small--i.e., the normal distribution has a very narrow, high peak that strongly indicates the presence of a target. The *MaM* estimator will decide in favor of the presence of a target if

$$1-q = f(0 \mid Z) < f(1 \mid Z) = q$$

i.e., if $q > \tfrac{1}{2}$. That is, the *MaM* estimator completely ignores critical information contained in the posterior--namely the variance σ^2. The *JoM* estimator, on the other hand, will decide in favor of the presence of a target if

$$1-q = f(\emptyset \mid Z) < c \cdot \underset{x}{argsup}\, f(x \mid Z) = \frac{cq}{\sqrt{2\pi}\,\sigma}$$

To evaluate this inequality we need to select a value for c. Assume that the prior multitarget density is uniform in the sense that $f_0(\emptyset \mid Z) = \tfrac{1}{2}$, $f_0(x \mid Z) = \tfrac{1}{2} A^{-1}$. If the prior is accepted as a standard, the balancing-point for deciding between yes-target and no-target is $\tfrac{1}{2} = f_0(0 \mid Z) = c \cdot argsup_x f_0(x \mid Z) = \tfrac{1}{2} c A^{-1}$ and thus $c = A$.

Accordingly, the test for the *JoM* estimator is

$$\frac{A}{\sigma} > \sqrt{2\pi}\,(q^{-1} - 1)$$

In other words, the *JoM* estimator makes a more nuanced decision by balancing the two pieces of information that the posterior contains (i.e., q and σ). Information supporting a no-target decision--e.g. q is small--can be countermanded by sufficiently strong information supporting a yes-target decision--i.e., σ is sufficiently small.

7.2 Analytical Comparison of the *MaM* and *JoM* Estimators. The purpose of this section is to derive an analytical comparision between the convergence behaviors of the *JoM* and *MaM* estimators in a simple model problem. Assume that a single motionless target in one dimension is located at x_0 and is observed by a single sensor whose noise density is $f(z \mid x) = f(z-x)$ where $f(x)$ is a mean-zero density. We also assume that the sensor detects the target with probability $p_D = 1$, and that the target is obscured by independent, uniformly distributed, and extremely dense Poisson clutter. That is, the typical observation has the form $Z = T \cup W$ where $T = T(X)$ is the single observation generated by the target and W are the clutter observations. Assume that the size of the region of interest is A. Multitarget states have the form $X = \{x\}$ and multitarget observations have the form $Z = \{z_1, \ldots, z_m\}$ or $Z = \emptyset$. The FISST multitarget calculus shows that the multitarget likelihoods for T, W are

$$f_T(\emptyset \mid x) = 0, \qquad f_T(z \mid x) = f(z-x), \qquad f_W(z_1 \quad ,z_m) = e^{-\lambda}\,\lambda^m A^{-m}$$

where λ is the Poisson parameter. The FISST calculus also shows that the true multitarget likelihood $f(Z \mid X)$ for all observations (target and clutter) is given by $f(z_1, \ldots, z_m \mid \emptyset) = e^{-\lambda}\,\lambda^m A^{-m}$, $f(\emptyset \mid \emptyset) = e^{-\lambda}$, and

$$f(z_1, \ldots, z_m \mid x) = f(z_1, \ldots, z_m \mid \emptyset)\left(\frac{A}{\lambda}\sum_{i=1}^{m} f(z_i - x)\right), \qquad f(\emptyset \mid x) \quad 0$$

Define $g(\emptyset \mid x) = 0$, $g(\emptyset \mid \emptyset) = 1$, and

$$g(z_1, \ldots, z_m \mid x) = \frac{A}{\lambda}\sum_{i=1}^{m} f(z_i - x), \qquad g(z_1, \ldots, z_m \mid \emptyset) = 1$$

Let $Z_1, \ldots, Z_k$ be a sequence of observation-sets and $Z^{(k)} = \{Z_1, \ldots, Z_k\}$. Also, let $f_0(\emptyset) = 1-Q$, $f_0(x) = QA^{-1}$ be the multitarget prior distribution. Since the posterior distribution is

$$f(X \mid Z^{(k)}) = \frac{f(Z_1 \mid X) \cdots f(Z_k \mid X)\, f_0(X)}{\int f(Z_1 \mid Y) \cdots f(Z_k \mid Y)\, f_0(Y)\,\delta Y} = \frac{g(Z_1 \mid X) \cdots g(Z_m \mid X)\, f_0(X)}{\int g(Z_1 \mid X) \cdots g(Z_m \mid X)\, f_0(Y)\,\delta Y}$$

it follows that we can use $g(Z \mid X)$ as though it were the actual multitarget likelihood function. Now since λ is assumed to be very large, if Z is any observation-set then with very high probability $Z - \{x_0\}$ is a large sample of data drawn from the uniform clutter distribution $c(z) = A^{-1}$. Consequently, the quantity

$$K(x) = \frac{1}{|Z|}\sum_{z \in Z} f(z - x)$$

is a Parzen kernel estimator of the uniform distribution: $K(x) \approx c(x) = A^{-1}$. So,

$$g(Z|x) \qquad \frac{|Z|}{\lambda} + \frac{A}{\lambda} f(x_0 - x)$$

Define

$$\alpha \qquad \frac{\lambda_{geo}}{\lambda} \approx \frac{(|Z_1| \cdots |Z_k|)^{1/k}}{\lambda}$$

where λ_{geo} is the geometric mean of the Poisson distribution and $\lambda_{geo} < \lambda$. Thus $\alpha < 1$

Case I: Target Not in the Scene. Suppose that a target is not present in the scene. Then $g(Z^{(k)} \mid \emptyset) = 1$, $g(Z^{(k)} \mid x) \approx \alpha^k$ and so $f(0 \mid Z^{(k)}) = (1-Q)/(1-Q+Q\alpha^k)$. If we choose the prior as the reference standard, then the *JoM* constant c is determined by the equation $1-Q = c \cdot Q \cdot A^{-1}$ or $c = A(Q^{-1}-1)$. Then "no-target" tests are:

$$Q^{-1}-1 > \alpha^k \quad (MaM), \qquad 1 > \alpha^k \quad (JoM)$$

Case II: Target Present in the Scene, with High-Resolution Sensor. By saying that the sensor resolution is very good we mean that $AH/\lambda \;\; 1$ where $H = f(0)$. In this case

$$g(Z^{(k)}|x) \qquad \left(\frac{|Z_1|}{\lambda} + \frac{A}{\lambda} f(x_0 - x)\right) \cdots \left(\frac{|Z_k|}{\lambda} + \frac{A}{\lambda} f(x_0 - x)\right) \approx \alpha^k + \frac{A^k}{\lambda^k} f(x_0 - x)^k$$

Assume that $f(x)$ is "Gaussian-like" in the sense that $f(0)^k = H^k$ and $\int f(x)^k\, dx \approx H^{k-1}/k^{1/2}$. That is, $f(x)^k$ has only negligible probability outside of the interval $[-H^1k^{1/2}, +H^1k^{1/2}]$. Then

$$f(X|Z^{(k)}) \propto \begin{cases} 1-Q & \text{if } X = \emptyset \\ QA^{-1}(\alpha^k + A^k\lambda^{-k} f(x_0 - x)) & \text{if } X = \{x\} \end{cases}$$

and it may be shown that

$$\int f(x|Z^{(k)})\, dx = Q\alpha^k + \frac{Q}{\lambda\sqrt{k}}\left(\frac{HA}{\lambda}\right)^{k-1}$$

and, consequently, that the "yes-target" tests for the *MaM* and *JoM* estimators are:

$$\lambda(Q^{-1}-2)\sqrt{k+1} < \left(\frac{HA}{\lambda}\right)^k \quad (MaM), \qquad 1 < \left(\frac{HA}{\lambda}\right)^k \quad (JoM)$$

Therefore, *MaM* will generally converge more slowly than *JoM*. However, the differences will be most pronounced when there is significant prior belief that no target is present in the scene (i.e., $Q \approx 0$).

8. Conclusions

In this paper we have demonstrated that when one tries to generalize recursive Bayesian nonlinear filtering to the *multitarget case with unknown number of targets*, one quickly encounters serious troubles that directly bear on

practice. Specifically, *Bayes-optimal multitarget estimation and filtering encounters fundamental conceptual and practical difficulties when the number of targets is unknown*. What is at issue is neither theoretical hair-splitting nor mere mathematical "bookkeeping." Rather, what is actually at stake is *our ability to do Bayes-optimal multitarget filtering and estimation at all and, moreover, our ability to even know what "Bayes-optimal" means in such a context*. We explained the following key points underlying multitarget recursive Bayesian nonlinear filtering:

- Reconceptualize a group of unknown targets, of unknown number, as a *single joint multitarget system*
- Reconceptualize target number and individual target states as just specific parameters of a *single joint multitarget state*
- Reconceptualize priors, posteriors, and likelihoods as *single densities involving an unknown number of targets and target states*
- Reconceptualize detection and estimation as different aspects of a *single, simultaneous statistical process*
- Recognize that one cannot just "declare victory": reconceptualization is the easy part, because it leads to unexpected theoretical and practical difficulties that cannot be swept under the rug

We concretely illustrated the meaning of these points by contrasting true multitarget recursive Bayesian nonlinear filtering with three approaches that fail to address these difficulties: the EAMLE and generalized EAMLE filters, and a renaming of a truncated version of FISST called "joint multitarget probabilities (JMP". We concluded that the *JoM* multitarget state estimator should exhibit better performance than the *MaM* estimator.

References

[1] Y. Bar-Shalom and X.-R. Li (1993) *Estimation and Tracking: Principles, Techniques, and Software*, Artech House

[2] Y. Bar-Shalom and T.E. Fortman (1988) *Tracking and Data Association*, Academic Press

[3] M. Bardin (1984) "Multidimensional Point Processes and Random Closed Sets," *J. Applied Probability*, vol. 21, pp. 173-178

[4] C.A. Barlow, L.D. Stone, M.D. Finn (1996) "Unified data fusion," *Proc. 9th Nat'l Symp. on Sensor Fusion*, pp. 321-330

[5] R.E. Bethel and G.J. Paras (1994) "A PDF multi-target tracker," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 30 No. 2, pp. 386-403

[6] I.R. Goodman, R.P.S. Mahler, and H.T. Nguyen (1997) *Mathematics of Data Fusion*, Kluwer

[7] Y.C. Ho and R.C.K. Lee (1964) "A Bayesian Approach to Problems in Stochastic Estimation and Control," *IEEE Trans. on Automatic Control*, Vol. AC-9, pp. 333-339

[8] J. Goutsias, R. Mahler, and H.T. Nguyen, eds. (1997) *Random Sets: Theory and Applications*, Springer, 1997

[9] A.H. Jazwinski (1970) *Stochastic Processes and Filtering Theory*, Academic Press

[10] E.W. Kamen (1992) "Multiple target tracking based on symmetric measurement equations," *IEEE Trans. on Auto. Contr.*, Vol. 37, pp. 371-374

[11] E.W. Kamen, Y.J. Lee, and C.R. Sastry (1994) "A parallel SME filter for tracking multiple targets in three dimensions," *SPIE Proceedings* Vol. 2235, pp. 417-428

[12] E.W. Kamen and C.R. Sastry (1993) "Multiple target tracking using products of position measurement," *IEEE Trans. on Aerospace and Electronics Systems*, Vol. 29, pp. 476-493

[13] K. Kastella (1996) "Discrimination Gain for Sensor Management in Multitarget Detection and Tracking," *Proc. 1996 IMACS Multiconference on Computational and Engineering Applications (CESA'96); Symposium on Control, Optimization, and Supervision, Vol. 1*, Lille France, July 9-12 1996, pp. 167-172

[14] K. Kastella (1995) "Event-Averaged Maximum Likelihood Estimation and Mean-Field Theory in Multitarget Tracking," *IEEE Trans. Auto. Control*, Vol. 40 No. 6, pp. 1070-1074

[15] K. Kastella (1997) "Information Theory and Sensor Management," in A. Friedman (ed.), *Mathematics in Industrial Problems, Part 9*, Springer-Verlag, pp. 36-44

[16] K. Kastella (1997) "Joint multitarget probabilities for detection and tracking," SPIE Vol. 3086, pp. 122-128

[17] K. Kastella (1993) "Theater Acoustic Correlator Tracker," Lockheed Martin internal technical report, May 10 1993

[18] V. Krishnamurthy and R.J. Evans (1995) "The data association problem for Hidden Markov Models," *Proc. 34th IEEE Conf. on Decision and Contr.*, Dec. 1995, New Orleans, pp. 2764-2765

[19] A.D. Lanterman, M.I. Miller, D.L. Snyder, and W.J. Miceli (1994) "Jump-diffusion processes for the automated understanding of FLIR scenes," *SPIE Proceedings*, Vol. 2234, pp. 416-427

[20] R. Mahler (1994) "A Bayesian Interpretation of the Random-Set Approach to Data Fusion," Lockheed Martin Internal

Technical Report, dated April 11 1994

[21] R. Mahler (1997) "Decisions and Data Fusion," *Proc. 1997 Nat'l Symp. on Sensor and Data Fusion*, April 14-17 1997, M.I.T. Lincoln Laboratories, pp. 71-87

[22] R. Mahler (1994) "Global Integrated Data Fusion," *Proc. Seventh Nat'l Symp. on Sensor Fusion*, Vol. I (Unclassified), Sandia National Laboratories, Albuquerque NM, March 16-18 1994, ERIM, Ann Arbor, pp. 187-199

[23] R. Mahler (1996) "Global Optimal Sensor Allocation," *Proceedings of the Ninth National Symposium on Sensor Fusion*, Vol. I (Unclassified), Mar. 12-14 1996, Naval Postgraduate School, Monterey CA, pp. 347-366

[24] R. Mahler (1998) "Global posterior densities for sensor management," SPIE Vol. 3365, pp. 252-263

[25] R. Mahler (1998) "Information for fusion management and performance estimation," SPIE Vol. 3374, pp. 64-75

[26] R. Mahler (1995) "Information Theory and Data Fusion," *Proc. Eight Nat'l Symp. on Sensor Fusion*, Vol. I (Unclassified), Dallas TX, March 15-17 1995, ERIM, Ann Arbor, pp. 279-292

[27] R. Mahler (1993) "Integrated Random-Set Estimation and Detection: Numerical Example," Lockheed Martin Internal Technical Report, dated January 25 1994

[28] R. Mahler (1997) "Measurement models for ambiguous evidence using conditional random sets," *SPIE Proceedings*, Vol. 3068, pp. 40-51

[29] R. Mahler (1998) "Multisource, multitarget filtering: A unified approach," *SPIE Proceedings*, Vol. 3373, pp. 296-307

[30] R. Mahler (1999) "Multitarget Markov Motion Models," *SPIE Proceedings*, Vol. 3720, to appear

[31] R. Mahler (1995) Nonadditive probability, finite-set statistics, and information fusion," *Proc. 34th IEEE Conf. on Decision and Control*, New Orleans, Dec. 1995, pp. 1947-1952

[32] R. Mahler (1994), "The random set approach to data fusion," SPIE Vol. 2234, pp. 287-295

[33] R. Mahler (1996) "Unified data fusion: fuzzy logic, evidence, and rules," SPIE Vol. 2755, pp. 226-237

[34] R. Mahler (1996/1994) "A Unified Foundation for Data Fusion," in F.A. Sadjadi, ed., *Selected Papers on Sensor and Data Fusion*, SPIE Vol. MS-124, SPIE Optical Engineering Press, Bellingham WA, pp. 325-345; reprinted from *Proc. Seventh Joint Service Data Fusion Symp.*, Vol. I Part 1 (Unclassified), Oct. 25-28 1994, Johns Hopkins Applied Physics Laboratory, Laurel MD, Naval Air Development Center, Warminster PA, pp. 153-173

[35] R. Mahler (1995) "Unified nonparametric data fusion," SPIE Vol. 2484, pp. 66-74

[36] G. Mathéron (1975) *Random Sets and Integral Geometry*, J. Wiley, New York

[37] M.I. Miller, R.S. Teichman, A. Srivistava, J.A. O'Sullivan, and D.L. Synder (1993) "Jump-diffusion processes for automated tracking-target recognition," *1993 SPIE Conf. on Information Sciences and Systems*, Baltimore, March 24-26, Johns Hopkins

[38] D.J. Muder and S.D. O'Neil (1993) "The multi-dimensional SME filter for multitarget tracking," *SPIE Proceedings* Vol. 1954, pp. 587-599

[39] N. Portenko, H. Salehi, and A. Skorokhod (1997) "On optimal filtering of multitarget tracking systems based on point processes observations," *Random Operators and Stochastic Equations*, vol. 1, pp. 1-34

[40] S. Musick, K. Kastella, and R. Mahler (1998) "A practical implementation of joint multitarget probabilities," SPIE Vol. 3374, pp. 26-37

[41] H.W. Sorenson (1988) "Recursive Estimation for Nonlinear Dynamic Systems," in J.C. Spall, ed., *Bayesian Analysis of Time Series and Dynamic Models*, Marcel Dekker

[42] A. Srivistava, N. Cutaia, M.I. Miller, J.A. O'Sullivan, and D.L. Snyder (1992) "Multi-target narrowband direction finding and tracking using motion dynamics," *Proc. 30th Allerton Conf. on Communication, Control, and Computing*, U. of Illinois, Urbana, pp. 563-570

[43] H.L. van Trees (1968) *Detection, Estimation, and Modulation Theory, Part I*, John Wiley and Sons

[44] A. Wald (1949) "Note on the consistency of the maximum likelihood extimator," *Ann. Math. Stat.*, vol. 20, pp. 595-601

[45] R.B. Washburn (1987) "A random point process approach to multiobject tracking," *Proc. Amer. Contr. Conf.*, Vol. 3, pp. 1846-1852

[46] X. Xie and R.J. Evans (1991) "Multiple target tracking and multiple frequency line tracking using Hidden Markov Models," *IEEE Trans. on Sig. Proc.*, Vol. 39, pp. 2659-2676

Appendix: "Generalized EAMLE" [17]

To illustrate the CMLE method, consider the problem of tracking a collection of 1-dimensional targets moving in a surveillance volume of length L. The number of targets T is known. The state of the system is completely specified by the T-vector $X = (x_1,...,x_T)^T$. From a Kalman prediction step or some other type of prior estimation, we assume Gaussian a priori density is available for x. The prior is $P_0(X) = N(X;X_0,\Sigma)$ where $N(X;Y,\sigma)$ denotes a normal distribution in X with mean Y and covariance matrix σ....We assume that the target positions are a scanned by sensor [sic] that has uniform clutter with density ρ. The known detection probability $P_d \leq 1$ is the same for all of the targets. For the valid detections, the sensor is Gaussian with variance R. A scan of data then consists of the set $Z = \{z_1,...,z_M\}$. The sensor provides no information about which of the z_m are clutter and which are target-originated. The CMLE consists of (1) computing the likelihood $P(Z \mid X)$ and then 2) using conventional maximum likelihood estimation (MLE) (or maximum a posteriori estimation (MAP)) to estimate X,

$$\hat{X} = arg \max_{X} (\ln P_0(X) + \ln P(Z|X)) \qquad (1)$$

To evaluate $P(Z \mid X)$ we average over all possible assignments of the M measurements in the scan Z to the T tracks X....Observe that $P(Z \mid X)$ is implicitly conditioned on the number of targets present T, but we suppress that for now. This will become important for the detection problem, however. [Using mean-field methods we can write]

$$P(Z|X) = \int \left(\prod_m d\eta_m\right)\left(\prod_d \zeta_t\right) \exp[E(\eta_m,\zeta_t,z_m,x_t)] \qquad (13)$$

...This integral can be approximately evaluated using the saddle point method common throughout statistical physics and quantum field theory..The dominant contribution to the integral given by the stationary points $\hat{\eta}_m$ and $\hat{\varsigma}_t$ with respect to η_m and ς_t...The integral is given by

$$P(Z|X) = c' \exp\left(E(\hat{\eta}_m,\hat{\zeta}_t,z_m,x_t)\right)$$

where c' is a normalization constant that depends on T....Combining $P(Z \mid X)$ with the prior X yields the log-likelihood function to be minimized:

$$\Lambda = \frac{1}{2}(X-X_0)^T\Sigma^{-1}(X-X_0) - E(\eta_m,\zeta_t,z_m,x_t)$$

To form the mean field coherent maximum likelihood position estimate (MFCMLE),we must simultaneously minimize Λ with respect to X, η_m, and ς_t ...

2.3.3 MFT and Detection. We now present our approach for extending the MFCMLE to attack the problem of acoustic detection in theater ASW using the problem of detecting 1-dimensional random walkers against a clutter background as an example. Suppose that we have K scans of data $Z^1,...,Z^K$. Each scan $Z^k = \{z_1^k,...,z_{M(k)}^k\}$ has $M(k)$ measurements in it where $M(k)$ is a random variable that depends on false alarm rate, the number of targets present and their detection probabilities. We now denote the full data set $Z = (Z^1,...,Z^K)$. As in the tracking case studied above, we assume that the detection probability P_d for the targets and the false alarm density ρ are known. To estimate the number of tracks, we compute $P(Z \mid T)$ where T is the number of tracks. The MLE estimate for T is then

$$\hat{T} = arg \max_T P(Z|T). \qquad (20)$$

In principal, once an estimate for T is generated, the location of the tracks is an independent process. However, we will see that evaluation of $P(Z \mid T)$ will involve saddle point evaluation of an integral and that this saddle point will yield X. Note that for this K scan problem, X is now a $T \times K$ matrix with elements x_t^k, the position of target t at time k. We define $X^k = (x_1^k,...,x_T^k)$ to be the vector of target locations at time k (analogous to the system state vector in the tracking examples above) and $X_t = \{x_t^1,...,x_t^K\}$ to be the trajectory of a single target. To compute $P(Z \mid T)$, we use Bayes rule to decompose the probability in terms of the T-track state densities $P(X \mid T)$,

$$P(Z|T) = \int dX\, P(Z|X)\, P(X|T), \qquad (21)$$

where $dX = \Pi_{k=1}^K \Pi_{t=1}^T dx_t^k$. In Eq. (21) $P(X \mid T)$ represents our a priori knowledge about target behavior. There are two salient features that characterize Brownian walkers. First, they arise with uniform density in the target region so that $P(x_t^1) = 1/L$, $t = 1,...,T$. Second, if at time k the target is at x_t^k then the density for its location at time $k+1$ is normal with mean x_t^k and variance Q, $P(x_t^{k+1} \mid x_t^k) = N(x^{k+1};x_t^k,Q)$. This exhausts our a priori knowledge of the walkers so we have $P(X \mid T) = \text{PI}_{t=1}^T (P(x_t^1)\, \text{PI}_{t=1}^{K-1} P(x_t^{k+1} \mid x_t^k))$. The other factor in the integrand of Eq. (21) is $P(Z \mid X) = \text{PI}_{k=1}^K P(Z^k \mid X^k)$ where $P(Z^k \mid X^k)$ is given by the integral Eqs. (13,15). Defining $d\varsigma = \Pi_{k=1}^K \Pi_{t=1}^T d\varsigma_t^k$ and $d\eta = \Pi_{k=1}^K \Pi_{m=1}^{M(k)} d\eta_m^k$, we can combine these factors to express $P(Z \mid T)$ as the integral

$$P(Z|T) = \tilde{c} \int dX\, d\eta\, d\zeta\, \exp\left[-\frac{1}{2Q}\sum_{t=1}^{T}\sum_{k=1}^{K-1}(x_t^{k+1}-x_t^k)^2 + \sum_{k=1}^{K} E^k(\eta_m^k,\zeta_t^k,z_m^k,x_t^k)\right], \qquad (22)$$

where $E(\eta_m^k,\varsigma_{kt},z_m^k,x_t^k)$ is given by Eq. (15). Now, this integral can also be solved using saddle point methods. The detection algorithm then consists of computing Eq. (22) for various values of T and using Eq. (20) to pick the best value of T. Observe that the saddle point evaluation then yields estimates for the tracks $\hat{X}$....One subtle issue that arises here is the evaluation of the normalization $\tilde{c}$. For the tracking problem, we locate the stationary point of the integrand. However in this case, we must compare the value of the integral for different values of T which entails at least approximate evaluation of $\tilde{c}$. $\tilde{c}$ can be expressed in terms of the determinant of a matrix whose dimension is given by the total number of variables in the problem, so some type of approximation scheme will probably be required. Also, note that Eq. (22) must be evaluated for each possible value of T. An approach to this problem is to define a new function using the approximation to the Kronecker δ-function of Eq. (8),

$$P(\tau) = \sum_T P(Z|T)\,\delta(\tau - T) = \lim_{a\to\infty}(2\pi a)^{-1/2}\int d\mu\, e^{-\mu^2/2a+i\mu T}\sum_{T=0}^{\infty} e^{-i\mu T}P(Z|T) \qquad (23)$$

$P(\tau)$ is now a function of a continuous variable. Many of the terms in the (22) can be combined and it is probably possible to evaluate the sum over T in (23) in closed form. This has the advantage the it replaces separate integrals by a single one.